Semantic Technologies for the Production and Publication of Open Data in ACI - Automobile Club d'Italia

Paolo Bouquet[d], Daniela Caltabiano[a], Emanuela Catoni[a], Antonio Fabrizi[a], Domenico Lembo[b], Mauro Minenna[a], Andrea Molinari [d], Flavio Pompermaier[d], Mario Punchina[a], Giacomo Ronconi[b,c], Marco Ruzzi[c], Valerio Santarelli[b,c], Federico Scafoglieri[b]

(a) ACI Informatica
⟨lastname⟩@informatica.aci.it

(b) Sapienza Università di Roma
⟨lastname⟩@diag.uniroma1.it

(c) OBDA Systems
⟨lastname⟩@obdasystems.com

(d) OKKAM
⟨lastname⟩@okkam.it

Semantic technologies combine knowledge representation techniques with artificial intelligence in order to achieve a more effective management of enterprise knowledge bases, thanks to the separation of the conceptual level of the applications from the logical and physical ones, and to the automatic reasoning services they deploy for data access and control. In this context, *Ontology-based Data Management* (OBDM) [3] has consolidated itself as a paradigm for data integration and governance, based on a three-tier architecture: the ontology, the data sources, and the *mappings*, which declaratively link the ontology predicates to the data in the sources.

In this talk[1] we present a joint project by Sapienza University of Rome, the Automobile Club d'Italia (ACI), and OKKAM S.r.l.[2], a spinoff of the University of Trento. The objectives of the project were the definition of an ontology of ACI's Public Vehicle Register (PRA) and car tax domains, the development of an OBDM system to access the data through such ontology, and the creation of a web portal for the publication of ACI's car parc data in *Linked Open* format.

The Ontology and Mappings. The ACI ontology was designed using Eddy[3], a graphical OWL ontology editing system developed by Sapienza and OBDA Systems[4], a spin-off of Sapienza. It consists of roughly 750 entities and 600 OWL 2 logical axioms, and is divided into 8 modules, reflecting the partition of ACI's domain into as many logical areas. Among the most important modules, the one that describes the technical characterization and classification of vehicles uses advanced modelling patterns to capture changes in vehicle properties over time; the statistical module focuses on statistical

---

[2] `www.okkam.it/`
[3] `www.obdasystems.com/eddy`
[4] `www.obdasystems.com`

Linked Open data published by ACI, and comprises elements of the *Data Cube ontology*[5], a W3C standard vocabulary for statistical multi-dimensional data; lastly, portions of the *GeoNames*[6] ontology have been incorporated into the model of the Italian territory. The mappings, which describe the semantic connections between the ontology and the data in ACI's information systems, were built using the Mastro Protégé[7] plug-in.

Data Access and Linked Open Data. The OBDM system processes SPARQL queries over the ontology, exploiting the Mastro [2] reasoner, which reformulates the query w.r.t. the ontology and the mappings, to produce a new query which encodes the intentional reasoning of the system, and is directly applicable to the source relational databases. The datasets that are extracted through this process or directly from the relational datasources are then semantically annotated with respect to the ontology and encoded in RDF. These datasets are distributed in 5-star Linked Open Data format: along with being in a structured, non-proprietary and W3C standard format (RDF), they are semantically annotated through the ontology, and linked with external datasets, specifically GeoNames and ISPRA[8]. To build these links, we have used OKKAM's *Entity Name System* [1] (ENS), which manages unique and persistent entity identifiers in distributed data systems. The ENS can identify the same resource among different data flows, and reconcile different references to it by assigning a unique global identifier.

The ACI Linked Open Data Portal. One of the main achievements of the project was the development of a web portal through which ACI displays the ontology and publishes its datasets[9]. The portal was developed through Mastro Studio[10], a tool developed by OBDA Systems which provides an environment for end users to inspect the ontology, both in graphical and OWL form, consult its documentation in *wiki* format (enriched with hypertext links to ease navigation), and access the published datasets, in CSV and RDF formats. Each dataset is provided with a description, a distribution license, and its downloadable files. Furthermore, each resource in the datasets is dereferenced for web navigation inside the system through a unique URI which corresponds to a page containing the description of the resource and all the RDF triples in which it is envolved. Mastro Studio is based on the DKAN CMS for Open Data, and includes DKAN's core modules and a suite of custom modules. To provide its semantic data access features, Mastro Studio relies on Mastro through a RESTful web service interface.

References

1. P. Bouquet and A. Molinari. A global entity name system (ENS) for data ecosystems. *PVLDB*, 6(11):1182–1183, 2013.
2. G. De Giacomo, D. Lembo, M. Lenzerini, A. Poggi, R. Rosati, M. Ruzzi, and D. F. Savo. Mastro: A reasoner for effective ontology-based data access. In *Proc. of ORE-2012*, volume 858 of *CEUR*, 2012.
3. M. Lenzerini. Managing data through the lens of an ontology. *AI Magazine*, 39(2):65–74, 2018.

---

[5] https://www.w3.org/TR/vocab-data-cube/

[6] http://www.geonames.org/ontology/documentation.html

[7] www.obdasystems.com/mastro-protege-plugin

[8] http://dati.isprambiente.it/

[9] lod.aci.it

[10] www.obdasystems.com/mastrostudio